

January 18, 2000

Comparison of Tests for the Cointegrating Rank of a VAR Process with a Structural Shift*

by

Helmut Lütkepohl

Humboldt University Berlin

Pentti Saikkonen

University of Helsinki

and

Carsten Trenkler

Humboldt University Berlin

Abstract

Two different types of tests for the cointegrating rank of VAR processes with a deterministic shift in the level have been proposed in the literature. The first proposal is based on the LR principle using a specific Gaussian model set-up. In the second proposal the time series are adjusted for deterministic terms first and then LR type tests are applied to the adjusted series. The local power of the two types of tests is derived and compared. Moreover, the small sample size and power properties of the tests are explored. It is found that the tests based on adjusted series generally have superior local power and size properties.

Key words: Local power, test size, cointegration, vector autoregressive process, error correction model

*We are grateful to the Deutsche Forschungsgemeinschaft, SFB 373, and the European Commission under the Training and Mobility of Researchers Programme (contract No. ERBFMRXCT980213), for financial support. The second author also thanks the Alexander von Humboldt Foundation for financial support under a Humboldt research award.

1 Introduction

Many economic time series include structural shifts of their levels caused by exogenous events that have occurred at a known time point during the observation period. The German unification is an example of such an event. It has caused shifts in macroeconomic time series such as Gross National Product (GNP) or measures of the money stock. Different tests have been proposed that can be used to test for the cointegrating rank of a system of variables in the presence of level shifts. In particular, Johansen & Nielsen (1993) (henceforth abbreviated as J&N) have proposed likelihood ratio (LR) tests for the cointegrating rank of a Gaussian vector autoregressive (VAR) process with a deterministic shift in the level and possibly other deterministic terms in the data generating process (DGP) such as mean terms, linear time trends and/or seasonal dummy variables. Alternative tests that allow for deterministic shifts in the levels and other deterministic terms have been considered by Saikkonen & Lütkepohl (2000a) (henceforth abbreviated as S&L). These authors suggest to estimate the deterministic terms in a first step, subtract them from the original series and then perform an LR type test on the adjusted series.

The purpose of this paper is to compare these two proposals for testing for the cointegrating rank of a VAR process with level shifts. The two alternative test statistics have quite different properties under the null hypothesis, as shown in the aforementioned articles. Whereas the null distribution of the Johansen-Nielsen statistics depend on the timing of the shift, the same is not true for the Saikkonen-Lütkepohl proposal. Given the differences in the properties of the tests, it is of interest to investigate their relative power. We will do so on the basis of local power derivations and small sample simulations.

A number of other studies deal with cointegration in the presence of structural shifts (e.g., Hansen (1992), Gregory & Hansen (1996), Campos, Ericsson & Hendry (1996), Quintos (1998), Seo (1998) and Inoue (1999)). All these studies differ from the present investigation in that alternative models are used and/or other aspects of a cointegration analysis are considered. As mentioned earlier, in the present study we focus on tests for the cointegrating rank of a VAR system with a deterministic shift in the level.

The structure of the paper is as follows. In the next section the DGP is specified, the assumptions underlying our analysis are laid out and the alternative cointegration tests are presented. Local power results are discussed in Section 3 and a simulation study exploring

the small sample properties of the tests is presented in Section 4. Conclusions are drawn in Section 5. The proofs of the local power results are provided in the Appendix.

The following general notation is used. The differencing operator is denoted by Δ , that is, for a time series or stochastic process y_t we define $\Delta y_t = y_t - y_{t-1}$. The symbol $I(d)$ denotes an integrated process of order d , that is, the stochastic part of the process is stationary or asymptotically stationary after differencing d times while it is still nonstationary after differencing just $d - 1$ times. Convergence in distribution or weak convergence is signified by \xrightarrow{d} and iid stands for independently, identically distributed. $O_p(\cdot)$ and $o_p(\cdot)$ are the usual symbols for convergence in probability. Moreover, $\|\cdot\|$ denotes the Euclidean norm. The trace and the rank of the matrix A are denoted by $\text{tr}(A)$ and $\text{rk}(A)$, respectively. If A is an $(n \times m)$ matrix of full column rank ($n > m$), we denote an orthogonal complement by A_\perp so that A_\perp is an $(n \times (n - m))$ matrix of full column rank and such that $A'A_\perp = 0$. The orthogonal complement of a nonsingular square matrix is zero and the orthogonal complement of a zero matrix is an identity matrix of suitable dimension. An $(n \times n)$ identity matrix is denoted by I_n . For matrices A_1, \dots, A_s , $\text{diag}(A_1 : \dots : A_s)$ is the block-diagonal matrix with A_1, \dots, A_s on the diagonal. LS, GLS and RR are used to abbreviate least squares, generalized least squares and reduced rank, respectively. LR is short for likelihood ratio, DGP stands for data generation process and VAR and ECM abbreviate vector autoregressive and error correction model, respectively. If the lower bound of the summation index of a sum exceeds the upper bound, the sum is defined to be zero.

2 The Model

Suppose an observed n -dimensional time series $y_t = (y_{1t}, \dots, y_{nt})'$ ($t = 1, \dots, T$) is generated by the following mechanism:

$$y_t = \mu_0 + \mu_1 t + \delta d_t + x_t, \quad t = 1, 2, \dots, \quad (2.1)$$

where μ_i ($i = 0, 1$) and δ are unknown $(n \times 1)$ parameter vectors and d_t is a step dummy variable defined as

$$d_t = \begin{cases} 0, & t < T_1 \\ 1, & t \geq T_1 \end{cases}. \quad (2.2)$$

This variable allows to take into account a sudden change in the mean of the process as it occurs, for instance, in German macroeconomic time series at the time of the reunification.

We assume that T_1 is known a priori and that $T_1 \leq T$. It is also convenient and not restrictive from a practical point of view to assume that $T_1 > p$ and, furthermore, that

$$\lim_{T \rightarrow \infty} \frac{T_1}{T} = a_1 \quad \text{with} \quad 0 < a_1 < 1. \quad (2.3)$$

In other words, the break point T_1 may be thought of as occurring at a fixed proportion of the full sample size if an asymptotic analysis is performed where $T \rightarrow \infty$. S&L also permit $a_1 = 1$ and, hence, the break may be viewed as having occurred a fixed number of periods before the end of the sample period in their framework. We do not consider this possibility here to simplify a comparison with the J&N tests.

Note that our assumption of a single break point and, hence, a single step dummy is also made to facilitate the theoretical analysis in the following. Generalizations to more than one step dummy and impulse dummies in addition are possible but will not be considered here for simplicity. A similar comment applies for seasonal dummies. It is also possible to exclude the trend term from the model, that is, $\mu_1 = 0$ may be assumed a priori. This case will be discussed later on.

The term x_t in (2.1) is an unobservable stochastic error which is assumed to be a VAR(p) process with ECM representation

$$\Delta x_t = \Pi x_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta x_{t-j} + \varepsilon_t, \quad t = 1, 2, \dots, \quad (2.4)$$

where Π and Γ_j ($j = 1, \dots, p-1$) are $(n \times n)$ matrices and ε_t is zero mean Gaussian white noise so that $\varepsilon_t \sim N(0, \Omega)$. Although less restrictive assumptions are possible for ε_t , assuming normality is convenient for the comparison of tests we intend to perform. The initial values x_{-p+1}, \dots, x_0 are assumed to be from some fixed probability distribution which does not depend on the sample size.

We assume that x_t is at most $I(1)$ and it is cointegrated with cointegrating rank r . This assumption implies in particular that the matrix Π can be written as

$$\Pi = \alpha \beta', \quad (2.5)$$

where α and β are $(n \times r)$ matrices of full column rank. It is well-known that $\beta' x_t$ and Δx_t are then zero mean (asymptotically) stationary processes. Moreover, defining $\Phi = I_n - \Gamma_1 - \dots - \Gamma_{p-1}$ and $C = \beta_\perp (\alpha'_\perp \Phi \beta_\perp)^{-1} \alpha'_\perp$, we have

$$x_t = C \sum_{j=1}^t \varepsilon_j + \xi_t, \quad t = 1, 2, \dots, \quad (2.6)$$

where ξ_t is a zero mean $I(0)$ process (see Johansen (1995, Chapter 4)).

We are interested in testing whether the assumption made for the rank of the matrix Π is correct. In other words, for some prespecified rank r_0 , we wish to test the null hypothesis

$$H_0(r_0) : \text{rk}(\Pi) = r_0 \quad \text{vs.} \quad H_1(r_0) : \text{rk}(\Pi) > r_0. \quad (2.7)$$

In order to present the tests to be compared in the following, it is useful to rewrite the model (2.1)/(2.4) as

$$\begin{aligned} \Delta y_t &= \nu + \alpha(\beta' y_{t-1} - \tau(t-1) - \theta d_{t-1}) + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j} + \varepsilon_t \\ &= \nu + \Pi^* y_{t-1}^* + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j} + \varepsilon_t, \\ t &= p+1, p+2, \dots, \end{aligned} \quad (2.8)$$

where $\nu = -\Pi\mu_0 + \Phi\mu_1$, $\tau = \beta'\mu_1$, $\theta = \beta'\delta$, $\Pi^* = \alpha[\beta' : \tau : \theta]$ and $y_{t-1}^{*'} = [y'_{t-1}, -(t-1), -d_{t-1}]$. Notice that here Δd_{t-j} is an impulse dummy which has the value one at $t = T_1 + j$ and is zero elsewhere. Moreover,

$$\gamma_j = \begin{cases} \delta, & j = 0 \\ -\Gamma_j \delta, & j = 1, \dots, p-1 \end{cases}. \quad (2.9)$$

Equation (2.8) specifies an ECM for the observed series y_t . J&N start from this model with any number of impulse dummies and no restrictions for the associated parameters. They derive the following LR statistic for our testing problem. For a sample y_1, \dots, y_T , define $z'_{t-1} = (1, \Delta y'_{t-1}, \dots, \Delta y'_{t-p+1}, \Delta d_t, \dots, \Delta d_{t-p+1})$ and

$$M_T = (T-p)^{-1} \left[\sum_{t=p+1}^T y_{t-1}^* y_{t-1}^{*'} - \sum_{t=p+1}^T y_{t-1}^* z'_{t-1} \left(\sum_{t=p+1}^T z_{t-1} z'_{t-1} \right)^{-1} \sum_{t=p+1}^T z_{t-1} y_{t-1}^{*'} \right].$$

Moreover, let $\hat{\Pi}^*$ be the LS estimator of the matrix Π^* in the model (2.8), denote the corresponding LS residuals by $\hat{\varepsilon}_t$ and define

$$\hat{\Omega} = (T-p)^{-1} \sum_{t=p+1}^T \hat{\varepsilon}_t \hat{\varepsilon}_t'. \quad (2.10)$$

Denoting by $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_n$ the ordered generalized eigenvalues obtained as solutions of

$$\det(\hat{\Pi}^* M_T \hat{\Pi}^{*'} - \lambda \hat{\Omega}) = 0, \quad (2.11)$$

the LR statistic for testing the pair of hypotheses (2.7) is given by

$$LR_{J\&N}(r_0) = (T - p) \sum_{j=r_0+1}^n \log(1 + \hat{\lambda}_j). \quad (2.12)$$

Note that this is the LR statistic under Gaussian assumptions if no restrictions are available for the parameters of the impulse dummies. Thus, it is not precisely the one corresponding to the model (2.1), because, if that model is the point of departure, the parameters of the impulse dummies are known to satisfy (2.9).

A similar statistic is obtained if there is no linear trend term in the model and, thus, $\mu_1 = 0$ is imposed a priori. In that case the model corresponding to (2.8) is

$$\Delta y_t = \Pi^0 y_{t-1}^0 + \sum_{j=1}^{p-1} \Gamma_j \Delta y_{t-j} + \sum_{j=0}^{p-1} \gamma_j \Delta d_{t-j} + \varepsilon_t, \quad t = p+1, p+2, \dots, \quad (2.13)$$

where $\Pi^0 = \alpha[\beta' : \eta : \theta]$ with $\eta = -\beta'\mu_0$ and $y_{t-1}^0 = [y'_{t-1}, 1, -d_{t-1}]'$. The resulting test statistic will be denoted by $LR_{J\&N}^0(r_0)$.

S&L use the model in (2.8) to obtain first stage estimators for the parameters of the error process x_t , that is, for α , β , Γ_j ($j = 1, \dots, p-1$) and Ω . In this case, a conventional RR regression cannot be used to obtain the ML estimators corresponding to the model (2.1)/(2.4) because there are nonlinear restrictions for the parameters in (2.8). To obtain the above mentioned first stage estimators these restrictions are ignored. Having obtained the first stage estimators, S&L present a GLS estimation procedure for the parameters of the deterministic part, $\tilde{\mu}_i$ ($i = 0, 1$) and $\tilde{\delta}$. When these estimators are available one can form a sample analog of the series x_t as $\tilde{x}_t = y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t - \tilde{\delta} d_t$ and use it to derive LM or LR type test statistics for the hypothesis $H_0(r_0)$ in the same way as in S&L. The LR type test statistic may be obtained as the usual LR test statistic from the feasible counterpart of the ECM (2.4), that is, it is determined from

$$\Delta \tilde{x}_t = \Pi \tilde{x}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta \tilde{x}_{t-j} + e_t, \quad t = p+1, \dots, T, \quad (2.14)$$

where e_t is an error term. The following general formulation discusses LR type test statistics because S&L found that LM type test statistics have inferior small sample properties.

The LR type statistic based on (2.14) is obtained in the usual way by solving the generalized eigenvalue problem $\det(\tilde{\Pi} \tilde{M}_T \tilde{\Pi}' - \lambda \tilde{\Omega}) = 0$, where $\tilde{\Pi}$ is the LS estimator of Π obtained

from (2.14), $\tilde{\Omega}$ is the corresponding residual covariance matrix and

$$\tilde{M}_T = (T-p)^{-1} \left[\sum_{t=p+1}^T \tilde{x}_{t-1} \tilde{x}_{t-1}' - \sum_{t=p+1}^T \tilde{x}_{t-1} \Delta \tilde{X}_{t-1}' \left(\sum_{t=p+1}^T \Delta \tilde{X}_{t-1} \Delta \tilde{X}_{t-1}' \right)^{-1} \sum_{t=p+1}^T \Delta \tilde{X}_{t-1} \tilde{x}_{t-1}' \right]$$

with $\Delta \tilde{X}_{t-1} = [\Delta \tilde{x}_{t-1}' : \cdots : \Delta \tilde{x}_{t-p+1}']'$. Denoting the resulting eigenvalues by $\tilde{\lambda}_1 \geq \cdots \geq \tilde{\lambda}_n$, the LR type statistic becomes

$$LR_{S\&L}(r_0) = (T-p) \sum_{j=r_0+1}^n \log(1 + \tilde{\lambda}_j). \quad (2.15)$$

The corresponding test statistic obtained by assuming $\mu_1 = 0$ a priori will be denoted by $LR_{S\&L}^0(r_0)$. In the next section we consider the local power properties of the tests based on $LR_{J\&N}(r_0)$, $LR_{S\&L}(r_0)$, $LR_{J\&N}^0(r_0)$ and $LR_{S\&L}^0(r_0)$.

3 Local Power Analysis

In this section we will give the limiting distributions of the tests presented in the previous section under local alternatives and we compare the resulting local power. We consider local alternatives of the form

$$H_T(r_0) : \Pi = \Pi_T = \alpha \beta' + T^{-1} \alpha_1 \beta_1', \quad (3.1)$$

where α and β are fixed $(n \times r_0)$ matrices of rank r_0 and α_1 and β_1 are fixed $(n \times (r - r_0))$ matrices of rank $r - r_0$ and such that the matrices $[\alpha : \alpha_1]$ and $[\beta : \beta_1]$ have full column rank r . In the following we use the assumptions from Johansen (1995) and Rahbek (1994) that the VAR order is $p = 1$ and the eigenvalues of the matrices $I_{r_0} + \beta' \alpha$ and $I_r + [\beta : \beta_1]' [\alpha : \alpha_1]$ are less than 1 in modulus. The next theorem gives the asymptotic distribution of the four test statistics considered in the foregoing under these assumptions. A proof is given in the Appendix.

Theorem 1

Let $\mathbf{W}(u)$ denote a Brownian motion with covariance matrix Ω and $\mathbf{K}(t)$ the Ornstein-Uhlenbeck process defined by the integral equation

$$\mathbf{K}(u) = \alpha_{\perp}' \mathbf{W}(u) + \alpha_{\perp}' \alpha_1 \beta_1' \beta_{\perp} (\alpha_{\perp}' \beta_{\perp})^{-1} \int_0^u \mathbf{K}(s) ds \quad (0 \leq u \leq 1). \quad (3.2)$$

Furthermore, define $\mathbf{N}(s) = (\alpha'_\perp \Omega \alpha_\perp)^{-1/2} \mathbf{K}(s)$, $\bar{\mathbf{N}}(s) = \mathbf{N}(s) - \int_0^1 \mathbf{N}(u) du$,

$$\mathbf{N}^0(s) = \begin{bmatrix} \mathbf{N}(s) \\ 1 \\ \mathbf{1}_{[a_1, 1]}(s) \end{bmatrix}, \quad \mathbf{N}^*(s) = \begin{bmatrix} \bar{\mathbf{N}}(s) \\ s - \frac{1}{2} \\ \mathbf{1}_{[a_1, 1]}(s) - (1 - a_1) \end{bmatrix},$$

$\mathbf{N}_*(s) = \mathbf{N}(s) - s\mathbf{N}(1)$ and $d\mathbf{N}_*(s) = d\mathbf{N}(s) - ds\mathbf{N}(1)$. With this notation, if $\Pi_T = \alpha\beta + T^{-1}\alpha_1\beta_1'$,

$$LR_{J\&N}^0(r_0) \xrightarrow{d} \text{tr} \left\{ \left(\int_0^1 \mathbf{N}^0(s) d\mathbf{N}(s)' \right)' \left(\int_0^1 \mathbf{N}^0(s) \mathbf{N}^0(s)' ds \right)^{-1} \left(\int_0^1 \mathbf{N}^0(s) d\mathbf{N}(s)' \right) \right\}, \quad (3.3)$$

$$LR_{J\&N}(r_0) \xrightarrow{d} \text{tr} \left\{ \left(\int_0^1 \mathbf{N}^*(s) d\mathbf{N}(s)' \right)' \left(\int_0^1 \mathbf{N}^*(s) \mathbf{N}^*(s)' ds \right)^{-1} \left(\int_0^1 \mathbf{N}^*(s) d\mathbf{N}(s)' \right) \right\}, \quad (3.4)$$

$$LR_{S\&L}^0(r_0) \xrightarrow{d} \text{tr} \left\{ \left(\int_0^1 \mathbf{N}(s) d\mathbf{N}(s)' \right)' \left(\int_0^1 \mathbf{N}(s) \mathbf{N}(s)' ds \right)^{-1} \left(\int_0^1 \mathbf{N}(s) d\mathbf{N}(s)' \right) \right\}, \quad (3.5)$$

$$LR_{S\&L}(r_0) \xrightarrow{d} \text{tr} \left\{ \left(\int_0^1 \mathbf{N}_*(s) d\mathbf{N}_*(s)' \right)' \left(\int_0^1 \mathbf{N}_*(s) \mathbf{N}_*(s)' ds \right)^{-1} \left(\int_0^1 \mathbf{N}_*(s) d\mathbf{N}_*(s)' \right) \right\}, \quad (3.6)$$

where $\int_0^1 \mathbf{N}_*(s) d\mathbf{N}_*(s)'$ is an abbreviation for

$$\int_0^1 \mathbf{N}(s) d\mathbf{N}(s)' - \int_0^1 \mathbf{N}(s) ds \mathbf{N}(1)' - \mathbf{N}(1) \int_0^1 s d\mathbf{N}(s)' + \frac{1}{2} \mathbf{N}(1) \mathbf{N}(1)'.$$

□

Interestingly, the limiting distributions for $LR_{S\&L}^0(r_0)$ and $LR_{S\&L}(r_0)$ in the present situation are the same as for the corresponding test statistics for the case without a shift (see Saikkonen & Lütkepohl (2000b) and Lütkepohl & Saikkonen (2000)). In particular, $LR_{S\&L}^0(r_0)$ has the same limiting distribution as in the case without any deterministic terms (e.g., Saikkonen & Lütkepohl (1999)). For all the statistics the null distributions are obtained by replacing the Ornstein-Uhlenbeck processes $\mathbf{N}(s)$ by standard Brownian motions $\mathbf{B}(s)$ with identity covariance matrix.

The theorem shows that the local power depends on the difference $n - r_0$ and not on the dimension n and r_0 separately. Moreover, it does not depend on the actual values of the mean, trend and shift parameters μ_0 , μ_1 and δ . Furthermore, it is straightforward to check that

$$\mathbf{N}(s) = \mathbf{B}(s) + ab' \int_0^s \mathbf{N}(u) du, \quad (3.7)$$

where $\mathbf{B}(s)$ is again an $(n - r_0)$ -dimensional standard Brownian motion and the quantities a and b are given by $a = (\alpha'_\perp \Omega \alpha_\perp)^{-1/2} \alpha'_\perp \alpha_1$ and $b = (\alpha'_\perp \Omega \alpha_\perp)^{1/2} (\beta'_\perp \alpha_\perp)^{-1} \beta'_\perp \beta_1$ (cf. Johansen (1995, pp. 207-208)). Hence, the limiting distributions depend on α , β , Ω , α_1 and β_1 only through a and b . This implies, for instance, for the case $r - r_0 = 1$, where α_1 and β_1 are just vectors, that the limiting distribution only depends on the two parameters $l^2 = a'ab'b$ and $d^2 = (b'a)^2/(a'ab'b)$ (see Johansen (1995, Corollary 14.5)). The parameter $l^2 = 0$ if and only if $H_0(r_0)$ is true. Hence, $l = \sqrt{l^2}$ may be viewed as the distance of the local alternative from the null hypothesis. The second parameter, d^2 , falls in the interval $(0, 1]$. The case $d^2 = 0$ corresponds to $I(2)$ processes which are not permitted by our assumptions. Still, it is useful to know that local alternatives with d values close to zero correspond to processes close to being $I(2)$.

We have simulated the local power resulting from the asymptotic distributions in the theorem. For this purpose, we consider the case where α_1 and β_1 are $(n \times 1)$ vectors and simulate the discrete time counterpart of the $(n - r_0)$ -dimensional Ornstein-Uhlenbeck process $\mathbf{N}(s)$ as $\Delta \mathbf{N}_t = \frac{1}{T} ab' \mathbf{N}_{t-1} + \epsilon_t$ ($t = 1, \dots, T = 1000$) with $\epsilon_t \sim \text{iid } N(0, I_{n-r_0})$, $\mathbf{N}_0 = 0$,

$$b' = \begin{cases} 1 & \text{for } n - r_0 = 1 \\ (1, 0) & \text{for } n - r_0 = 2 \\ (1, 0, 0) & \text{for } n - r_0 = 3 \end{cases}$$

and

$$a' = \begin{cases} -\sqrt{l^2} & \text{for } n - r_0 = 1 \\ \left(-\sqrt{l^2 d^2}, \sqrt{l^2(1 - d^2)} \right) & \text{for } n - r_0 = 2 \\ \left(-\sqrt{l^2 d^2}, \sqrt{l^2(1 - d^2)}, 0 \right) & \text{for } n - r_0 = 3 \end{cases}.$$

From the generated \mathbf{N}_t we compute

$$A_T = \frac{1}{T^2} \sum_{t=1}^T F_t F'_t \quad \text{and} \quad B_T = \frac{1}{T} \sum_{t=1}^T F_t \Delta \mathbf{N}'_t,$$

with

$$F_t = \begin{bmatrix} \mathbf{N}_{t-1} \\ 1 \\ d_{t-1} \end{bmatrix} \quad \text{for } LR_{J\&N}^0(r_0),$$

$$F_t = \begin{bmatrix} \mathbf{N}_{t-1} - T^{-1} \sum_{t=1}^T \mathbf{N}_{t-1} \\ t - 1 - \frac{1}{2}(T - 1) \\ d_{t-1} - (T - T_1 - 1)/T \end{bmatrix} \quad \text{for } LR_{J\&N}(r_0)$$

and $F_t = \mathbf{N}_{t-1}$ for $LR_{S\&L}^0(r_0)$. The limiting distributions of the three statistics are then simulated as $\text{tr}(B_T' A_T^{-1} B_T)$. Furthermore, using

$$A_T = \frac{1}{T^2} \sum_{t=1}^T \left[\sum_{k=1}^{t-1} (\Delta \mathbf{N}_k - \overline{\Delta \mathbf{N}}) \right] \left[\sum_{k=1}^{t-1} (\Delta \mathbf{N}_k - \overline{\Delta \mathbf{N}}) \right]'$$

and

$$B_T = \frac{1}{T} \sum_{t=1}^T \left[\sum_{k=1}^{t-1} (\Delta \mathbf{N}_k - \overline{\Delta \mathbf{N}}) \right] (\Delta \mathbf{N}_t - \overline{\Delta \mathbf{N}})',$$

with $\overline{\Delta \mathbf{N}} = T^{-1} \sum_{t=1}^T \Delta \mathbf{N}_t$ and considering again $\text{tr}(B_T' A_T^{-1} B_T)$ gives the limiting distribution of $LR_{S\&L}(r_0)$. The resulting rejection frequencies for the cases $n - r_0 = 1$, $n - r_0 = 2$ and $n - r_0 = 3$ for different values of d and l are plotted in Figures 1 - 3. In the figures the local power of the tests which allow for a structural shift is also compared to tests which do not allow for a shift in the deterministic term. The tests denoted as LR_{Joh} and LR_{Joh}^0 are the appropriate LR tests presented by Johansen (1995) for processes without a shift. The limiting distributions of LR_{Joh} and LR_{Joh}^0 are given in Saikkonen & Lütkepohl (1999). To simulate these distributions, F_t is adjusted accordingly by deleting the component associated to the shift dummy.

$LR_{S\&L}^0$ clearly outperforms the corresponding Johansen type tests assuming no linear trend ($LR_{J\&N}^0$ and LR_{Joh}^0), especially in case of $n - r_0 = 1$ (see Figure 1, Panel A). For a range of values of l , the rejection probabilities of $LR_{S\&L}^0$ are more than twice as large as those of $LR_{J\&N}^0$ and LR_{Joh}^0 . It is also obvious that for the Johansen type tests including a shift term results in a loss in local power. This loss tends to be slightly larger for greater values of d , that is, for processes well away from being $I(2)$. It is remarkable, however, that the $LR_{S\&L}^0$ tests which include a shift term have uniformly better local power in Figures 1 - 3, Panels A than LR_{Joh}^0 which does not allow for such a term.

Comparing the tests which make allowance for a linear trend it can be seen that the $LR_{S\&L}$ tests again have a local power advantage for most of the combinations of l and d . In particular, the power advantage tends to increase with d . In other words, it is larger for processes away from the $I(2)$ region than for processes with parameters close to the $I(2)$

region. For large values of l , that is, for local alternatives far away from the null hypothesis, the local power of LR_{Joh} is about the same as and in some cases even slightly larger than the one of $LR_{S\&L}$. Note, however, that a similar result for $LR_{J\&N}$ is only obtained with even higher values of l , because accommodating a shift in the presence of a linear trend reduces the local power of the Johansen type tests, as in the case without a linear trend term.

There are also a couple of other observations worth noting in Figures 1 - 3. First, the local power of the tests assuming $\mu_1 = 0$ is generally better than the local power of the tests allowing for a linear trend and sometimes the difference is quite substantial. Thus, it is worth excluding the trend term if it is not needed in the DGP. Another important issue is the dependence of the power on $n - r_0$, the number of stochastic trends under $H_0(r_0)$. Figures 1 - 3 show that increasing $n - r_0$ results in a loss of power for all tests. For instance, this means that a single cointegration relation is more difficult to detect in a three-dimensional than in a bivariate process. A similar result was also obtained by Johansen (1995) and Saikkonen & Lütkepohl (1999) for processes without a level shift.

Summarizing the results, it can be concluded that in the case without a trend the S&L tests are clearly superior. In general their power advantage is quite dramatic. In the presence of a linear trend they are still superior in large parts of the parameter space but there are also cases where the J&N tests have more local power. Of course, local power properties are asymptotic properties and, hence, they are informative about the performance of the tests in large samples when alternatives close to the null hypothesis are of interest. In small samples the situation may be different. This issue is investigated in the following section.

4 Small Sample Comparison of Tests

We have compared the properties of the tests in a small Monte Carlo experiment. The simulations are based on the following process x_t from Toda (1994, 1995) which was also used by some other authors:

$$x_t = \begin{bmatrix} \Psi & 0 \\ 0 & I_{n-r} \end{bmatrix} x_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \text{iid } N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} I_r & \Theta \\ \Theta' & I_{n-r} \end{bmatrix} \right), \quad (4.1)$$

where $\Psi = \text{diag}(\psi_1, \dots, \psi_r)$ is an $(r \times r)$ diagonal matrix and Θ is $(r \times (n - r))$. Specifically, we will consider bivariate and four-dimensional processes. For instance, in the bivariate case, if $r = 0$, Ψ and Θ vanish and the process consists of two nonstationary components.

If the cointegrating rank is 1, $\Psi = \psi_1$ with $|\psi_1| < 1$. In that case $\Theta = \theta$ is a scalar which represents the instantaneous correlation between the two components. Since the test results are invariant to the parameter values of the deterministic component we use $\mu_i = 0$ ($i = 0, 1$) and $\delta = 0$ throughout. In other words, the deterministic part is actually zero. Thereby we can compare the performance of our tests with other tests which do not allow for shifts.

Samples of sizes 100 and 200 plus 50 presample values starting with initial values of zero were generated. The number of replications is 10,000. The rejection frequencies given in Tables 1 - 3 and Figures 4 - 8 are based on asymptotic critical values for a test level of 5%. The rejection frequencies are not corrected for the actual small sample sizes because such corrections will not be available in practice.

For a given set of parameter values and a given sample size, the results for the test statistics are based on the same generated time series. Hence, the entries in the tables and the figures are not independent but can be compared directly. Still, for judging the results, it may be worth taking into account that the standard error of an estimator of a true rejection probability P based on 10,000 replications of the experiment is $s_P = \sqrt{P(1-P)/10,000}$. Hence, for example, $s_{0.05} = 0.0022$. Thus, the two-standard-error confidence interval for $P = 0.05$ is $[0.0456, 0.0544]$. It is also important to note that in the simulations the tests were not performed sequentially. Thus, the results for testing $H_0(1) : \text{rk}(\Pi) = 1$ are not conditioned on the outcome of the test of $H_0(0) : \text{rk}(\Pi) = 0$ etc..

In Table 1 and Figure 4 the performance of the tests LR_{Joh} , LR_{Joh}^0 , LR_{ta} and LR_{ta}^0 which do not allow for a shift is also presented. The latter two tests were proposed by Saikkonen & Lütkepohl (2000b) for processes without a shift. They have the same asymptotic distribution as $LR_{S\&L}^0$ and $LR_{S\&L}$, respectively. Of course, the small sample properties may still be different. Therefore, they are included in the comparison. In Table 1 the true cointegrating rank is zero so that $\varepsilon_t \sim \text{iid } N(0, I_2)$ and, hence, the components of ε_t are independent. Here the results for testing $H_0(0) : \text{rk}(\Pi) = 0$ give an indication of the actual sizes of the tests for a nominal size of 5%.

It is seen that the location of the break point (T_1) does not matter much for the size of the tests in the present situation. In most cases the differences in the sizes for different values of T_1 are smaller than twice the standard error $s_{0.05}$. This result was also obtained for other processes with different cointegrating ranks and instantaneous residual correlation for which

Table 1. Relative Rejection Frequencies of Tests for Bivariate DGPs with Cointegrating Rank $r = 0$, VAR Order $p = 1$, Sample Size $T = 100$, Nominal Significance Level 0.05.

Assumed Break Point	Test Statistic	Rank under H_0		Test Statistic	Rank under H_0	
		$r_0 = 0$	$r_0 = 1$		$r_0 = 0$	$r_0 = 1$
none	LR_{Joh}	0.061	0.003	LR_{Joh}^0	0.058	0.005
	LR_{ta}	0.047	0.004	LR_{ta}^0	0.054	0.015
$T_1 = 25$	$LR_{J\&N}$	0.070	0.005	$LR_{J\&N}^0$	0.063	0.004
	$LR_{S\&L}$	0.046	0.007	$LR_{S\&L}^0$	0.056	0.019
$T_1 = 50$	$LR_{J\&N}$	0.068	0.004	$LR_{J\&N}^0$	0.064	0.006
	$LR_{S\&L}$	0.046	0.006	$LR_{S\&L}^0$	0.053	0.016
$T_1 = 75$	$LR_{J\&N}$	0.064	0.003	$LR_{J\&N}^0$	0.063	0.005
	$LR_{S\&L}$	0.048	0.006	$LR_{S\&L}^0$	0.051	0.016

the results are not reported here. Moreover, a similar observation was made for the power of the tests. Therefore, the subsequent comparison focuses on a break point $T_1/T = 0.75$.

Obviously, whether or not a shift is accommodated does not matter much for the sizes of the tests. The same holds for the power of the S&L tests displayed in Figure 4, Panels A and C. A somewhat different result is obtained for the tests proposed by Johansen & Nielsen. The graphs in Panels B and D of Figure 4 show that their power can be more than 10 percentage points lower than the power of the corresponding tests not allowing for shifts in the levels.

All tests tend to be very conservative if the rank is overstated in the null hypothesis (see $r_0 = 1$ in Table 1 or $r_0 = 1$ if $r = 0$ in Table 2). The J&N tests are also quite conservative in case of processes with a true cointegrating rank $r = 1$ when there is no innovation correlation ($\theta = 0$). The size is larger, however, if $\theta = 0.8$. This outcome has been explained by Toda (1995). He shows that for a true cointegrating rank $r = 1$ the tests reject the null hypothesis of no cointegration more easily for $|\theta|$ closer to 1.

The performance of the $LR_{S\&L}$ test which allows for a deterministic trend also depends on θ , but in the opposite way. $LR_{S\&L}$ is conservative for $\theta = 0.8$ and has a more accurate size in case of no instantaneous innovation correlation. The $LR_{S\&L}^0$ test which assumes $\mu_1 = 0$ has roughly a correct size for both types of error correlations. Obviously, this latter test has

Table 2. Relative Rejection Frequencies of Tests for Bivariate DGPs with Cointegrating Rank $r = 0$ or 1 ($\psi_1 < 1$), VAR Order $p = 1$, Sample Size $T = 100$, Break Point $T_1 = 75$, Nominal Significance Level 0.05.

Test	$r = 0$		$r = 1$		
			$\psi_1 = 0.9$	$\psi_1 = 0.8$	$\psi_1 = 0.7$
Statistic	$r_0 = 0^*$	$r_0 = 1^*$	$r_0 = 1^*$	$r_0 = 1^*$	$r_0 = 1^*$
$\theta = 0$					
$LR_{J\&N}$	0.064	0.003	0.005	0.019	0.036
$LR_{S\&L}$	0.048	0.006	0.025	0.048	0.056
$LR_{J\&N}^0$	0.063	0.005	0.007	0.024	0.038
$LR_{S\&L}^0$	0.051	0.016	0.047	0.055	0.049
$\theta = 0.8$					
$LR_{J\&N}$	0.064	0.003	0.032	0.059	0.064
$LR_{S\&L}$	0.048	0.006	0.020	0.028	0.035
$LR_{J\&N}^0$	0.063	0.005	0.044	0.062	0.061
$LR_{S\&L}^0$	0.051	0.016	0.052	0.048	0.049

* rank specified in the null hypothesis.

the most favorable size properties. Almost all size values lie inside the two-standard-error confidence interval mentioned above. In contrast, the size of $LR_{J\&N}^0$ strongly depends on the value of ψ_1 . Hence, in terms of size, $LR_{S\&L}^0$ is preferable. We mention, however, that in this respect the performance of the tests generally improves for sample size $T = 200$, as expected. The actual results are not shown to economize on space.

The power of the tests in the presently considered bivariate case with $T = 100$ is compared in Figure 5, where the true cointegrating rank is 1 (except for $\psi_1 = 1$). It can be seen that corresponding tests have similar power. None of the tests is uniformly superior. Not surprisingly, the tests assuming no linear trend have a higher power than the corresponding ones allowing for a trend. So it is worth applying the former group of tests if it is known that there is no linear trend in the data. The power of the tests for samples with $T = 200$ is presented in Figure 6. Although the power increases with the sample size, even $T = 200$ does not ensure a high power for all the values of θ and ψ_1 considered in the simulations. The general results found for the case $T = 100$ remain unchanged for $T = 200$.

Table 3. Relative Rejection Frequencies of Tests for Four-dimensional DGPs with Cointegrating Rank $r = 0$ or 1 ($\psi_1 < 1$), VAR Order $p = 2$, Sample Size $T = 100$, Break Point $T_1 = 75$, Nominal Significance Level 0.05 .

Test	$r = 0$		$r = 1$		
			$\psi_1 = 0.9$	$\psi_1 = 0.8$	$\psi_1 = 0.7$
Statistic	$r_0 = 0^*$	$r_0 = 1^*$	$r_0 = 1^*$	$r_0 = 1^*$	$r_0 = 1^*$
$\Theta = (0, 0, 0)$					
$LR_{J\&N}$	0.147	0.014	0.015	0.024	0.036
$LR_{S\&L}$	0.083	0.010	0.014	0.023	0.034
$LR_{J\&N}^0$	0.136	0.015	0.016	0.028	0.043
$LR_{S\&L}^0$	0.098	0.022	0.028	0.039	0.052
$\Theta = (0.4, 0.4, 0.8)$					
$LR_{J\&N}$	0.147	0.014	0.187	0.157	0.142
$LR_{S\&L}$	0.083	0.010	0.065	0.060	0.061
$LR_{J\&N}^0$	0.136	0.015	0.174	0.147	0.131
$LR_{S\&L}^0$	0.098	0.022	0.126	0.096	0.086

* rank specified in the null hypothesis.

We have also simulated four-dimensional systems to investigate the impact of the number of variables on the performance of the tests. Some results for processes with cointegrating ranks 0 and 1 are given in Table 3 and Figure 7. The results are obtained by estimating a VAR model of order $p = 2$ although the actual DGP in (4.1) has order one only. Thus, the estimated processes involve many more parameters than the bivariate processes. Thereby we intend to check the reaction of the tests to increased estimation uncertainty. In the absence of innovation correlation ($\Theta = 0$) all tests have a tendency to be conservative when testing $r_0 = 1$. In case of high innovation correlation the J&N tests display an excessive size distortion and reject far too often. This kind of size distortion is a typical problem of the Johansen type tests that emerges in large systems even if no shift term is considered as Gonzalo & Pitarakis (1999) have pointed out. The S&L tests also show size distortions and reject too often, although to a lesser extent than the J&N tests. Overall both S&L tests have more favorable size properties than the corresponding J&N tests.

In Figure 7 the power of the tests is depicted. Taking into account the excessive overre-

jection of the J&N tests, it turns out that the J&N and S&L tests, perform fairly similarly because the power curves of corresponding tests are almost parallel. Again, the tests assuming no deterministic trend have a slightly larger power than the test versions allowing for a trend, as expected.

In Figure 8, power results for four-dimensional processes with cointegrating rank $r = 2$ are presented using again a VAR order of $p = 2$. Obviously, under the present scenario, the power of all the tests is quite low. To demonstrate that this result is not a consequence of allowing for a deterministic shift, we also present power curves for the corresponding tests based on the assumption that there is no structural break. Clearly, these results indicate that in some situations finding cointegration ranks greater than one is not very likely. Again the J&N and S&L tests perform similarly in this respect.

Summarizing the results, it can be seen that the tests perform rather similarly with respect to power although each of the test proposals has relative advantages in specific situations. Generally, the S&L tests have more favorable size properties, however. Therefore, we recommend to use them in applied work.

5 Conclusions

In this study we have compared two types of tests for the cointegrating rank of a VAR process with a deterministic shift in the level at some known point in time. The first class of tests was derived by J&N using the LR principle under Gaussian assumptions. The second group of tests proceeds by estimating and subtracting all deterministic terms in a first step and then applying LR techniques to the adjusted series. These tests were proposed by S&L. Both classes of tests are based on very similar models and can be applied in the same situations. Therefore the question arises which of them should be used in practice.

A conceptual advantage of the S&L tests is that their asymptotic properties do not depend on where the shift actually occurs. Therefore, the same critical values can be used independent of the break date. The same is not true for the J&N tests whose limiting null distribution depends on the break date. Thus, if for instance new data become available, new critical values have to be determined. Usually the critical values have to be simulated for each specific situation whereas a single table with critical values is sufficient for the S&L tests.

Moreover, a detailed analysis of the local power properties of the two alternative classes of tests reveals that the S&L tests are generally superior in this respect as well. In particular, if no trend term is necessary in the DGP, the local power advantage of the S&L tests can be quite substantial. On the other hand, if a trend term is included in the DGP, there are also situations where both types of tests have quite similar local power or where the J&N tests are slightly superior.

We have also compared the two types of tests in small samples using a simulation study. It turns out that the size properties of the J&N tests are in some situations quite unsatisfactory. In some cases they are conservative and reject far too rarely whereas for other processes they overreject considerably. Although there are also cases where the size properties of the S&L tests are not quite satisfactory, they are generally much better in this respect than the J&N tests. On the other hand, the two classes of tests are quite similar in terms of small sample power when the differences in the sizes are accounted for. Therefore, overall the S&L tests are recommended for applied work.

Appendix A. Proof of Theorem 1

In proving Theorem 1, we draw on results and derivations given in Saikkonen & Lütkepohl (1999) and Lütkepohl & Saikkonen (2000). For simplicity we abbreviate these references as SL and L&S, respectively. In particular in the latter article only sketches of proofs are given. More details may be found in the discussion paper version which is available upon request. We will also repeatedly refer to Saikkonen & Lütkepohl (2000a) which we continue to abbreviate as S&L.

We shall first prove (3.3). The test statistic $LR_{J\&N}^0(r_0)$ is obtained from the model

$$\Delta y_t = \Pi^0 y_{t-1}^0 + \delta \Delta d_t + e_t^0, \quad t = 2, \dots, T,$$

where $y_{t-1}^0 = [y'_{t-1}, 1, -d_{t-1}]'$, $\Pi^0 = \alpha [\beta' : \eta : \theta] = \alpha \beta^{0'}$, and $e_t^0 = \varepsilon_t + T^{-1} \alpha_1 \beta_1^{0'} y_{t-1}^0$ with $\beta_1^0 = [\beta_1' : \eta_1 : \theta_1]'$. Furthermore, $\eta = -\beta' \mu_0$, $\theta = \beta' \delta$, $\eta_1 = -\beta_1' \mu_0$ and $\theta_1 = \beta_1' \delta$. Using a linear reparameterization we can write this model as

$$\Delta y_t = \alpha \underline{\beta}^{0'} \underline{x}_{t-1}^0 + \delta \Delta d_t + e_t^0, \quad t = 2, \dots, T,$$

where $\underline{x}_{t-1}^0 = [x'_{t-1}, T^{1/2}, -T^{1/2} d_{t-1}]'$ and $\underline{\beta}^0 = [\beta' : \underline{\eta} : \underline{\theta}]'$ with $\underline{\eta} = T^{-1/2} (\beta' \mu_0 + \eta)$ and $\underline{\theta} = T^{-1/2} (\theta - \beta' \delta)$. Moreover, by similar arguments we can write $e_t^0 = \varepsilon_t + T^{-1} \alpha_1 \underline{\beta}_1^{0'} \underline{x}_{t-1}^0$,

where $\underline{\beta}_1^0 = [\beta_1' : \underline{\eta}_1 : \underline{\theta}_1]'$ with $\underline{\eta}_1 = T^{-1/2}(\beta_1'\mu_0 + \eta_1)$ and $\underline{\theta}_1 = T^{-1/2}(\theta_1 - \beta_1'\delta)$. Note that the true values of the parameters $\underline{\eta}$, $\underline{\theta}$, $\underline{\eta}_1$ and $\underline{\theta}_1$ are zero.

Let R_{0t}^0 , R_{1t}^0 and E_t^0 be the LS residuals from a regression of Δy_t , \underline{x}_{t-1}^0 and e_t^0 on Δd_t , respectively, so that $R_{0t}^0 = \Delta y_t - \Delta y_t \Delta d_t$ and similarly for R_{1t}^0 and E_t^0 . Then consider the auxiliary RR regression model

$$R_{0t}^0 = \alpha \underline{\beta}^{0'} R_{1t}^0 + E_t^0. \quad (A.1)$$

From the above derivation it is apparent that, instead of the original model, the test statistic $LR_{J\&N}^0(r_0)$ can also be obtained from this infeasible model which can therefore be used for theoretical purposes. In particular, we can use this model and a general result in SL to derive the limiting distribution of the test statistic $LR_{J\&N}^0(r_0)$. To this end, we need to show that model (A.1) satisfies Assumption 1 of SL. The counterparts of X_t , ε_t , B , B_1 and $\Sigma_{\varepsilon\varepsilon}$ in that assumption are R_{1t}^0 , $\varepsilon_t - \varepsilon_t \Delta d_t$, $\underline{\beta}^0$, $\underline{\beta}_1^0$ and Ω , respectively, and, since the true value of $\underline{\beta}^0$ is $[\beta' : 0 : 0]'$, we can choose $\underline{\beta}_\perp^0 = \text{diag}[\beta_\perp : 1 : 1]$. For convenience, we define $R_{\varepsilon t}^0 = \varepsilon_t - \varepsilon_t \Delta d_t$.

It is easy to see that, as far as the verification of Assumption 1 of SL is concerned, R_{1t}^0 and $R_{\varepsilon t}^0$ can be replaced by \underline{x}_{t-1}^0 and ε_t , respectively. Thus, since \underline{x}_{t-1}^0 equals x_{t-1}^* in (A.19) of SL augmented by $-T^{1/2}d_{t-1}$ and since $\underline{\beta}^{0'} \underline{x}_{t-1}^0 = \beta' x_{t-1}$, we can proceed in the same way as on page 73 of that paper and show that $T^{-1} \sum_{t=1}^T \underline{\beta}_\perp^{0'} R_{1t}^0 R_{1t}^{0'} \underline{\beta}_\perp^0$ converges in probability to a positive definite constant matrix. Observing that

$$\underline{\beta}_\perp^{0'} R_{1t}^0 = \begin{bmatrix} \beta_\perp' x_{t-1} - \beta_\perp' x_{t-1} \Delta d_t \\ T^{1/2} - T^{1/2} \Delta d_t \\ -T^{1/2} d_{t-1} + T^{1/2} d_{t-1} \Delta d_t \end{bmatrix},$$

we can similarly show that $T^{-1} \sum_{t=1}^T \underline{\beta}_\perp^{0'} R_{1t}^0 R_{1t}^{0'} \underline{\beta}_\perp^0 = O_p(1)$. Thus, we have verified Assumptions 1(i) and (ii) of SL.

Next note that, by Theorem 14.1 of Johansen (1995), $T^{-1/2}(\beta_\perp' x_{[Ts]-1} - \beta_\perp' x_{[Ts]-1} \Delta d_{[Ts]}) \xrightarrow{d} \mathbf{K}(s)$. Thus, since $d_{[Ts]-1} \rightarrow \mathbf{1}_{[a_1, 1]}(s)$, the preceding expression of $\underline{\beta}_\perp^{0'} R_{1t}^0$ and a straightforward application of the continuous mapping theorem yield

$$T^{-2} \sum_{t=1}^T \underline{\beta}_\perp^{0'} R_{1t}^0 R_{1t}^{0'} \underline{\beta}_\perp^0 \xrightarrow{d} \int_0^1 \mathbf{K}^0(s) \mathbf{K}^0(s)' ds, \quad (A.2)$$

where $\mathbf{K}^0(s) = [\mathbf{K}(s)', 1, \mathbf{1}_{[a_1, 1]}(s)]'$. Because the r.h.s. of (A.2) corresponds to $B_\perp' G B_\perp$ in Assumption 1(iii) of SL and since we have verified Assumptions 1(i) and (ii) of that paper

we can conclude that Assumption 1(iii) holds. To verify Assumption 1(iv) we need to show that $T^{-1/2} \sum_{t=1}^T R_{\varepsilon t}^0 R_{1t}^{0'} \underline{\beta}^0 = O_p(1)$. This result can again be obtained in the same way as in SL (p. 73).

To prove Assumption 1(v) notice that

$$\alpha_{\perp}' T^{-1} \sum_{t=1}^T R_{\varepsilon t}^0 R_{1t}^{0'} \underline{\beta}_{\perp}^0 = \alpha_{\perp}' T^{-1} \sum_{t=1}^T \varepsilon_t x_{t-1}^{0'} \underline{\beta}_{\perp}^0 + o_p(1) \xrightarrow{d} \int_0^1 d\mathbf{W}(s) \mathbf{K}^0(s)', \quad (A.3)$$

where the last relation can be justified in the same way as in (A.21) of SL. Thus, Assumption 1(v) of that paper follows. Finally, since the present counterpart of ε_t in SL is $R_{\varepsilon t}^0$, Assumption 1(vi) is obvious, that is, $T^{-1} \sum_{t=1}^T R_{\varepsilon t}^0 R_{\varepsilon t}^{0'} = \Omega + O_p(T^{-1/2})$.

Since Assumption 1 of SL has been verified we can derive the limiting distribution of the test statistic $LR_{J\&N}^0(r_0)$ from Theorem 1 of that paper. In the same way as in SL (p. 73) we have $\underline{\beta}_{\perp}^{0'} \underline{\beta}_{\perp}^0 = \text{diag}[\beta_{\perp}' \alpha_{\perp} : 1 : 1]$ and $\underline{\beta}_{\perp}^{0'} \underline{\beta}_1^0 = [\beta_1' \beta_{\perp} : 0 : 0]'$. Since α_{\perp} is the counterpart of the matrix A_{\perp} in SL we can conclude from (A.2) and (A.3) that the counterpart of the matrix $(FB_{\perp}' GB_{\perp} + A_{\perp} S)'$ becomes

$$\int_0^1 \mathbf{K}^0(s) d\mathbf{W}(s)' \alpha_{\perp} + \int_0^1 \mathbf{K}^0(s) \mathbf{K}(s)' ds H' = \int_0^1 \mathbf{K}^0(s) d\mathbf{K}(s)',$$

where $H = \alpha_{\perp}' \alpha_1 \beta_1' \beta_{\perp} (\alpha_{\perp}' \beta_{\perp})^{-1}$ and the last equality is due to the definition of the process $\mathbf{K}(s)$. Hence, combining this and (A.2) with Theorem 1 of SL gives

$$LR_{J\&N}^0(r_0) \xrightarrow{d} \text{tr} \left\{ (\alpha_{\perp}' \Omega \alpha_{\perp})^{-1} \left(\int_0^1 \mathbf{K}^0(s) d\mathbf{K}(s)' \right)' \left(\int_0^1 \mathbf{K}^0(s) \mathbf{K}^0(s)' ds \right)^{-1} \left(\int_0^1 \mathbf{K}^0(s) d\mathbf{K}(s)' \right) \right\},$$

where $\alpha_{\perp}' \Omega \alpha_{\perp}$ is the counterpart of the matrix $A_{\perp}' \Sigma_{\varepsilon\varepsilon} A_{\perp}$ in the above mentioned theorem. Since it is straightforward to check that the above limiting distribution agrees with that in (3.3) we have proven the first assertion of Theorem 1.

Next consider the test statistic $LR_{J\&N}(r_0)$ which is obtained from the model

$$\Delta y_t = \nu + \Pi^* y_{t-1}^* + \delta \Delta d_t + e_t^*, \quad t = 2, \dots, T,$$

where $y_{t-1}^* = [y_{t-1}', -(t-1), -d_{t-1}]'$, $\Pi^* = \alpha [\beta' : \tau : \theta]' = \alpha \beta^{*'}$, and $e_t^* = \varepsilon_t + T^{-1} \alpha_1 \beta_1^{*'} y_{t-1}^*$ with $\beta_1^* = [\beta_1' : \tau_1 : \theta_1]'$. Furthermore, $\tau = \beta_1' \mu_1$ and $\tau_1 = \beta_1' \mu_1$, while $\theta = \beta_1' \delta$ and $\theta_1 = \beta_1' \delta$ as before. Using a linear reparameterization we can write this model as

$$\Delta y_t = \mu_1 + \alpha \underline{\beta}^{*'} \underline{x}_{t-1}^* + \delta \Delta d_t + \underline{e}_t^*, \quad t = 2, \dots, T,$$

where $\underline{x}_{t-1}^* = [x'_{t-1}, -T^{-1/2}(t-1), -T^{1/2}d_{t-1}]'$ and $\underline{\beta}^* = [\beta' : \underline{\tau} : \underline{\theta}]'$ with $\underline{\tau} = T^{1/2}(\tau - \beta'\mu_1)$. Further, $\underline{e}_t^* = \varepsilon_t + T^{-1}\alpha_1\beta_1^{*'}\underline{x}_{t-1}^*$ with $\underline{\beta}_1^* = [\beta_1' : \underline{\tau}_1 : \underline{\theta}_1]'$ and $\underline{\tau}_1 = T^{1/2}(\tau_1 - \beta_1'\mu_1)$. As before, $\underline{\theta} = T^{-1/2}(\theta - \beta'\delta)$ and $\underline{\theta}_1 = T^{-1/2}(\theta_1 - \beta_1'\delta)$. The true values of these two parameters are zero and so are those of $\underline{\tau}$ and $\underline{\tau}_1$.

Let R_{0t}^* , R_{1t}^* and E_t^* be the LS residuals from a regression of Δy_t , \underline{x}_{t-1}^* and \underline{e}_t^* on $[1, \Delta d_t]'$, respectively, and consider the auxiliary RR regression model

$$R_{0t}^* = \alpha \underline{\beta}^{*'} R_{1t}^* + E_t^*. \quad (A.4)$$

Test statistic $LR_{J\&N}(r_0)$ can also be obtained from this infeasible model and its limiting distribution can again be derived from Theorem 1 of SL. Thus, we have to show that model (A.4) satisfies Assumption 1 of that paper. Denoting by $R_{\varepsilon t}^*$ the LS residual of ε_t on $[1, \Delta d_t]'$, the counterparts of X_t , ε_t , B , B_1 and $\Sigma_{\varepsilon\varepsilon}$ in Assumption 1 of LS are R_{1t}^* , $R_{\varepsilon t}^*$, $\underline{\beta}^*$, $\underline{\beta}_1^*$ and Ω , respectively. Also, since the true values of the parameters τ , $\underline{\tau}_1$, $\underline{\theta}$ and $\underline{\theta}_1$ are zero, we can choose $\underline{\beta}_\perp^* = \text{diag}[\beta_\perp : 1 : 1]$.

It is again easy to see that the inclusion of the impulse dummy Δd_t in the model has no effect on asymptotic derivations. Thus, in the same way as in SL (p. 74) we can verify Assumptions 1(i) and (ii) or that $T^{-1} \sum_{t=1}^T \underline{\beta}^{*'} R_{1t}^* R_{1t}^{*'} \underline{\beta}^*$ converges in probability to a positive definite constant matrix and $T^{-1} \sum_{t=1}^T \underline{\beta}_\perp^{*'} R_{1t}^* R_{1t}^{*'} \underline{\beta}^* = O_p(1)$. Also, arguments similar to those used to obtain (A.2) show that

$$T^{-2} \sum_{t=1}^T \underline{\beta}_\perp^{*'} R_{1t}^* R_{1t}^{*'} \underline{\beta}_\perp^* \xrightarrow{d} \int_0^1 \mathbf{K}^*(s) \mathbf{K}^*(s)' ds, \quad (A.5)$$

where $\mathbf{K}^*(s) = [\overline{\mathbf{K}}(s)', s - \frac{1}{2}, \mathbf{1}_{[a_1, 1]}(s) - (1 - a_1)]'$ and $\overline{\mathbf{K}}(s)$ is the demeaned version of $\mathbf{K}(s)$. In the same way as in the sequel of (A.2) we can thus conclude that Assumption 1(iii) of SL holds with the matrix $B_\perp' G B_\perp$ given by the weak limit in (A.5). Moreover, using definitions it is straightforward to check that $T^{-1/2} \sum_{t=1}^T R_{\varepsilon t}^* R_{1t}^{*'} \underline{\beta}^* = O_p(1)$. Thus, Assumption 1(iv) of SL also holds.

Next consider the quantity

$$T^{-1} \sum_{t=1}^T R_{\varepsilon t}^* R_{1t}^{*'} \underline{\beta}_\perp^* = T^{-1} \sum_{t=1}^T \varepsilon_t R_{1t}^{*'} \underline{\beta}_\perp^* \xrightarrow{d} \int_0^1 d\mathbf{W}(s) \mathbf{K}^*(s)'. \quad (A.6)$$

Here the equality follows from definitions while the stated weak convergence is a straightforward extension of (A.26) of SL. Thus, we have shown that Assumption 1(v) of SL holds

whereas Assumption 1(vi) is obvious because, by standard arguments, $T^{-1} \sum_{t=1}^T R_{\varepsilon t}^* R_{\varepsilon t}^{*'} = T^{-1} \sum_{t=1}^T \varepsilon_t \varepsilon_t' + O_p(T^{-1/2}) = \Omega + O_p(T^{-1/2})$. Hence, we have verified Assumption 1 of SL. Since $\underline{\beta}_\perp^* = \underline{\beta}_\perp^0$ and $\underline{\beta}_1^* = \underline{\beta}_1^0$ we can repeat the arguments that follow (A.3) and find that the counterpart of the matrix $(FB'_\perp GB_\perp + A_\perp S)'$ in Theorem 1 of SL becomes

$$\int_0^1 \mathbf{K}^*(s) d\mathbf{W}(s)' \alpha_\perp + \int_0^1 \mathbf{K}^*(s) \mathbf{K}(s)' ds H' = \int_0^1 \mathbf{K}^*(s) d\mathbf{K}(s)'. \quad (\text{A.7})$$

In the same way as in the case of the test statistic $LR_{J\&N}^0(r_0)$ we can finally conclude from (A.5) and (A.7) that the limiting distribution of $LR_{J\&N}(r_0)$ is as stated in (3.4).

We shall next prove (3.6) and after that note that (3.5) can be proven along similar lines. Thus, consider $LR_{S\&L}(r_0)$ which can be obtained from the auxiliary RR regression model

$$\Delta \tilde{x}_t = \alpha \beta' \tilde{x}_{t-1} + \tilde{e}_t, \quad t = 2, \dots, T,$$

where

$$\begin{aligned} \tilde{x}_t &= y_t - \tilde{\mu}_0 - \tilde{\mu}_1 t - \tilde{\delta} d_{t-1} \\ &= x_t - (\tilde{\mu}_0 - \mu_0) - (\tilde{\mu}_1 - \mu_1) t - (\tilde{\delta} - \delta) d_t \end{aligned}$$

and

$$\begin{aligned} \tilde{e}_t &= \varepsilon_t + \alpha \beta' (\tilde{\mu}_0 - \mu_0) + \alpha \beta' (\tilde{\mu}_1 - \mu_1) (t - 1) + \alpha \beta' (\tilde{\delta} - \delta) d_{t-1} - (\tilde{\mu}_1 - \mu_1) \\ &\quad - (\tilde{\delta} - \delta) \Delta d_t + T^{-1} \alpha_1 \beta_1' (\tilde{\mu}_0 - \mu_0) + T^{-1} \alpha_1 \beta_1' (\tilde{\mu}_1 - \mu_1) (t - 1) \\ &\quad + T^{-1} \alpha_1 \beta_1' (\tilde{\delta} - \delta) d_{t-1} + T^{-1} \alpha_1 \beta_1' \tilde{x}_{t-1} \\ &\stackrel{def}{=} s_t + T^{-1} \alpha_1 \beta_1' \tilde{x}_{t-1}. \end{aligned}$$

The counterparts of X_t , ε_t , B , B_1 and $\Sigma_{\varepsilon\varepsilon}$ in Assumption 1 of SL are now \tilde{x}_{t-1} , s_t , β , β_1 and Ω , respectively.

The first step in the proof is to demonstrate that the nuisance parameter estimators have the following properties:

$$\beta' (\tilde{\mu}_0 - \mu_0) = O_p(T^{-1/2}) \quad (\text{A.8})$$

$$\beta'_\perp (\tilde{\mu}_0 - \mu_0) = O_p(1) \quad (\text{A.9})$$

$$\beta' (\tilde{\delta} - \delta) = O_p\left((T - T_1)^{-1/2}\right) \quad (\text{A.10})$$

$$\beta'_\perp (\tilde{\delta} - \delta) = O_p(1) \quad (\text{A.11})$$

$$\beta'(\tilde{\mu}_1 - \mu_1) = O_p(T^{-3/2}) \quad (A.12)$$

$$T^{1/2}\beta'_\perp(\tilde{\mu}_1 - \mu_1) \xrightarrow{d} \mathbf{K}(1), \quad (A.13)$$

where all the quantities converge weakly and jointly upon appropriate standardizations. Under the null hypothesis these results are proven in S&L. Under local alternatives, (A.8), (A.9), (A.12) and (A.13) are obtained in L&S in a model without dummy variables for related estimators. To demonstrate that (A.8) - (A.13) hold we first note that the consistency results given for the RR estimators of the parameters β , α and Ω in Lemma 3.1 of S&L also hold in the present context. To see this, the error term ε_t in S&L has to be replaced by $\varepsilon_t + T^{-1}\alpha_1\beta'_1x_{t-1}$ after which the arguments used in the previous proof in the case $a_1 < 1$ can be repeated. Since local alternatives are considered, the limit results in Johansen (1995, Chapter 14) involving weak convergence to an Ornstein-Uhlenbeck process instead of a Brownian motion have to be employed. This, however, does not alter any rates of convergence so that the previous consistency results will be obtained for the RR estimators of β , α and Ω . Thus, we have $\tilde{\alpha} = \alpha + O_p(T^{-1/2})$, $\tilde{\beta} = \beta + O_p(T^{-1})$ and $\tilde{\Omega} = \Omega + O_p(T^{-1/2})$, where appropriate normalizations, not affecting our test statistics, are assumed for the parameters α and β and their estimators.

To demonstrate that (A.8) - (A.12) hold, we can proceed in the same way as in the proof of Theorem 3.1 of S&L with the previous error term ε_t in the expression (A.13) replaced by $\varepsilon_t + T^{-1}\alpha_1\beta'_1x_{t-1}$. Thus, using the consistency results obtained above and the limit results in Johansen (1995, Chapter 14) we can first show that (A.14) and (A.15) of S&L still hold and that the appropriately standardized moment matrix of the auxiliary regression model (A.12) of S&L behaves asymptotically in the same way as before. From these facts it follows that (A.16) of S&L still holds and therefore (A.8) - (A.12) of the present paper are obtained in the same way as previously. Thus, we have to establish (A.13). In S&L it was concluded that, under the null hypothesis, the derivation of the limiting distribution of $\beta'_\perp\tilde{\mu}_1$ reduces to that considered in Saikkonen & Lütkepohl (2000b). It is straightforward to check that this conclusion also holds in the present context. After this, adjusting the arguments given in Saikkonen & Lütkepohl (2000b) in conjunction with the limit results in Johansen (1995, Chapter 14) shows that

$$\begin{aligned} T^{1/2}\beta'_\perp(\tilde{\mu}_1 - \mu_1) &= \beta'_\perp CT^{-1/2} \sum_{t=1}^T (\varepsilon_t + T^{-1}\alpha_1\beta'_1x_{t-1}) + o_p(1) \\ &= \beta'_\perp CT^{-1/2} \sum_{t=1}^T \Delta x_t + o_p(1), \end{aligned} \quad (A.14)$$

where the latter equality is based on the definition of the matrix C . Since we can assume that $\beta'_\perp \alpha_\perp = \beta'_\perp \beta_\perp$ (see SL, p. 72) we have $\beta'_\perp C = \alpha'_\perp$ and (A.13) follows from (A.14) and Theorem 14.1 of Johansen (1995).

Next we have to verify Assumption 1 of SL. The first condition of this assumption states that the matrix $T^{-1} \sum_{t=1}^T \beta' \tilde{x}_{t-1} \tilde{x}'_{t-1} \beta$ converges in probability to a positive definite matrix. Using the definition of \tilde{x}_t and (A.8), (A.10) and (A.12) it can be seen that the inclusion of the dummy variable d_t has no effect in this context so that the desired result is obtained from L&S. Using (A.8) - (A.13) it can similarly be seen that $T^{-1} \sum_{t=1}^T \beta' \tilde{x}_{t-1} \tilde{x}'_{t-1} \beta_\perp = O_p(1)$ so that Assumption 1(ii) of SL holds. Since it is further straightforward to check that the dummy variable d_t has no effect on the asymptotic behavior of the matrix $T^{-2} \sum_{t=1}^T \tilde{x}_{t-1} \tilde{x}'_{t-1}$ the arguments given in L&S show that Assumption 1(iii) of SL holds and also that

$$T^{-2} \sum_{t=1}^T \beta'_\perp \tilde{x}_{t-1} \tilde{x}'_{t-1} \beta_\perp \xrightarrow{d} \int_0^1 \mathbf{K}_*(s) \mathbf{K}_*(s)' ds, \quad (\text{A.15})$$

where $\mathbf{K}_*(s) = \mathbf{K}(s) - s\mathbf{K}(1)$. This is the counterpart of the matrix $B'_\perp G B_\perp$ in SL.

Regarding Assumption 1(iv) of SL, we have to show that $T^{-1/2} \sum_{t=1}^T s_t \tilde{x}'_{t-1} \beta = O_p(1)$. It is not difficult to see that as far as this result is concerned the effect of the dummy variable d_t is not different from the effect of the intercept term. Therefore, the desired result can be obtained from the arguments used in L&S. Since the same is true for Assumption 1(v) of SL we can in particular conclude that

$$\alpha'_\perp T^{-1} \sum_{t=1}^T s_t \tilde{x}'_{t-1} \beta_\perp \xrightarrow{d} \alpha'_\perp \int_0^1 d\mathbf{W}(s) \mathbf{K}_*(s)' - \mathbf{K}(1) \int_0^1 \mathbf{K}_*(s)' ds + H\mathbf{K}(1) \int_0^1 s \mathbf{K}_*(s)' ds. \quad (\text{A.16})$$

This term is the counterpart of $A'_\perp S$ in Assumption 1(v) of SL. Finally, as in L&S, it can be shown that $T^{-1} \sum_{t=1}^T s_t s'_t = \Omega + O_p(T^{-1/2})$ so that Assumption 1(vi) of SL holds. Using this fact, (A.15) and (A.16) it can be seen that the limiting distribution of the test statistic $LR_{S\&L}(r_0)$ is the same as in L&S. This proves (3.6). Omitting the trend term from the model and repeating the above arguments with appropriate modifications one can similarly prove (3.5). This completes the proof of Theorem 1.

References

- Campos, J., N.R. Ericsson & D.F. Hendry (1996), Cointegration tests in the presence of structural breaks, *Journal of Econometrics*, 70, 187 - 220.

- Gonzalo, J. & J.-Y. Pitarakis (1999), Dimensionality effects in cointegration analysis, in: R.F. Engle & H. White (eds.), *Cointegration, Causality, and Forecasting: A Festschrift in Honour of Clive W.J. Granger*, Oxford: Oxford University Press, pp. 212 - 229.
- Gregory, A.W. & B.E. Hansen (1996), Residual-based tests for cointegration in models with regime shifts, *Journal of Econometrics*, 70, 99 - 126.
- Hansen, B.E. (1992), Tests for parameter instability in regressions with $I(1)$ processes, *Journal of Business & Economic Statistics*, 10, 321 - 335.
- Inoue, A. (1999), Tests of cointegrating rank with a trend-break, *Journal of Econometrics*, 90, 215 - 237.
- Johansen, S. (1995), *Likelihood Based Inference in Cointegrated Vector Autoregressive Models*, Oxford: Oxford University Press.
- Johansen, S. & B. Nielsen (1993), Manual for the simulation program DisCo. Institute of Mathematical Statistics, University of Copenhagen.
- Lütkepohl, H. & P. Saikkonen (2000), Testing for the cointegrating rank of a VAR process with a time trend, *Journal of Econometrics*, 95, 177 - 198.
- Quintos, C.E. (1998), Stability tests in error correction models, *Journal of Econometrics*, 82, 289 - 315.
- Rahbek, A.C. (1994), The power of some multivariate cointegration tests, Discussion Paper.
- Saikkonen, P. & H. Lütkepohl (1999), Local power of likelihood ratio tests for the cointegrating rank of a VAR process, *Econometric Theory*, 15, 50 - 78.
- Saikkonen, P. & H. Lütkepohl (2000a), Testing for the cointegrating rank of a VAR process with structural shifts, *Journal of Business & Economic Statistics*, forthcoming.
- Saikkonen, P. & H. Lütkepohl (2000b), Trend adjustment prior to testing for the cointegrating rank of a VAR process, *Journal of Time Series Analysis*, forthcoming.
- Seo, B. (1998), Tests for structural change in cointegrated systems, *Econometric Theory*, 14, 222 - 259.

Toda, H.Y. (1994), Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present, *Review of Economics and Statistics*, 76, 66 - 79.

Toda, H.Y. (1995), Finite sample performance of likelihood ratio tests for cointegration ranks in vector autoregressions, *Econometric Theory*, 11, 1015-1032.

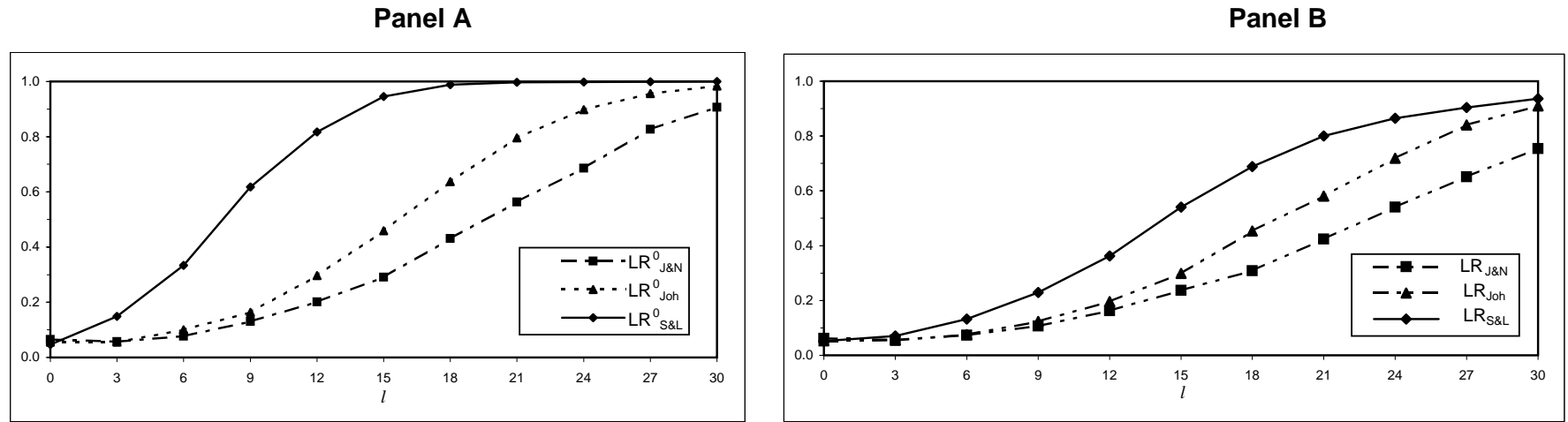


Figure 1. Local power of LR type tests for $n - r_0 = 1$.

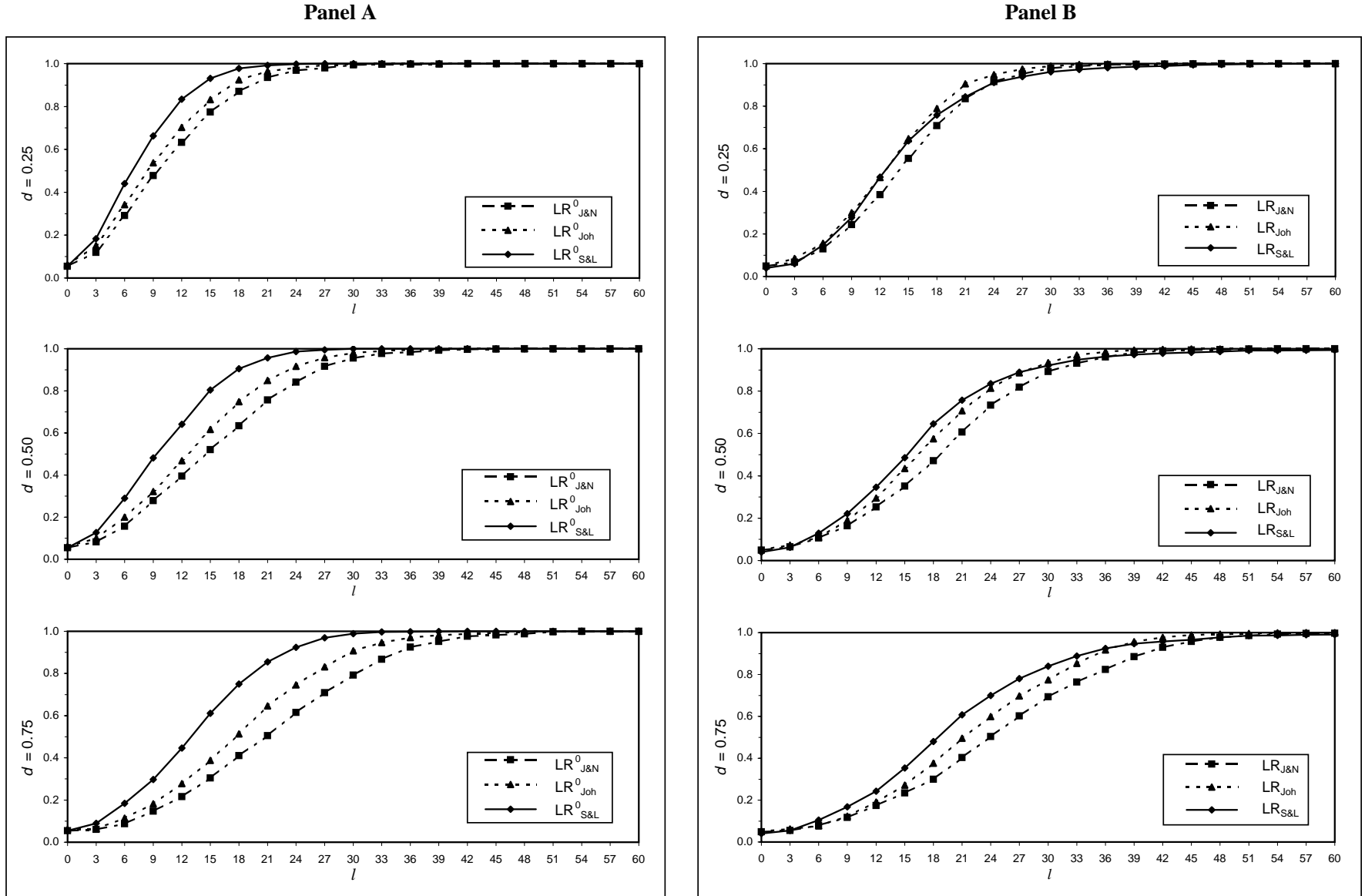


Figure 2. Local power of LR type tests for $n - r_0 = 2$.

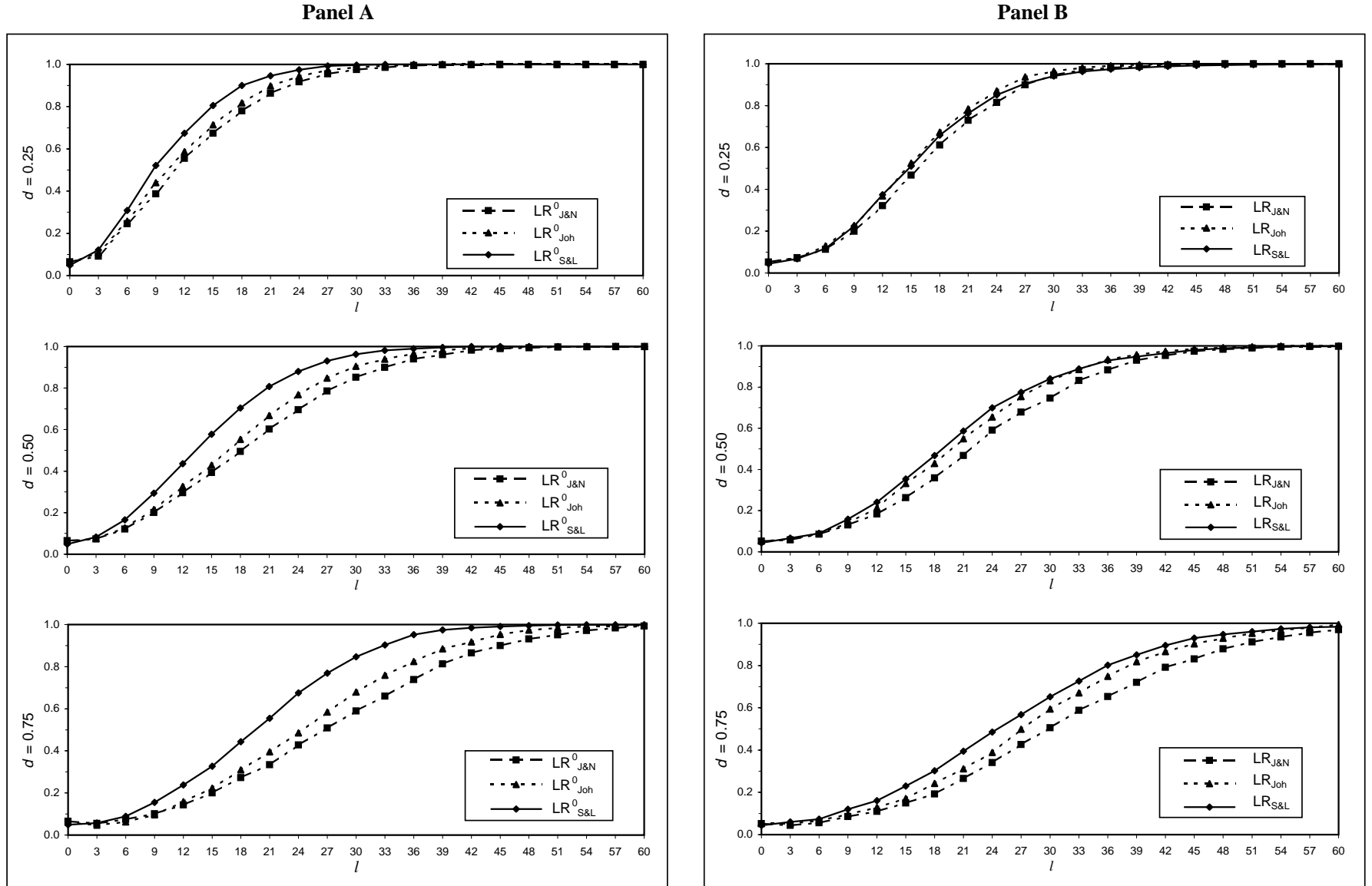


Figure 3. Local power of LR type tests for $n - r_0 = 3$

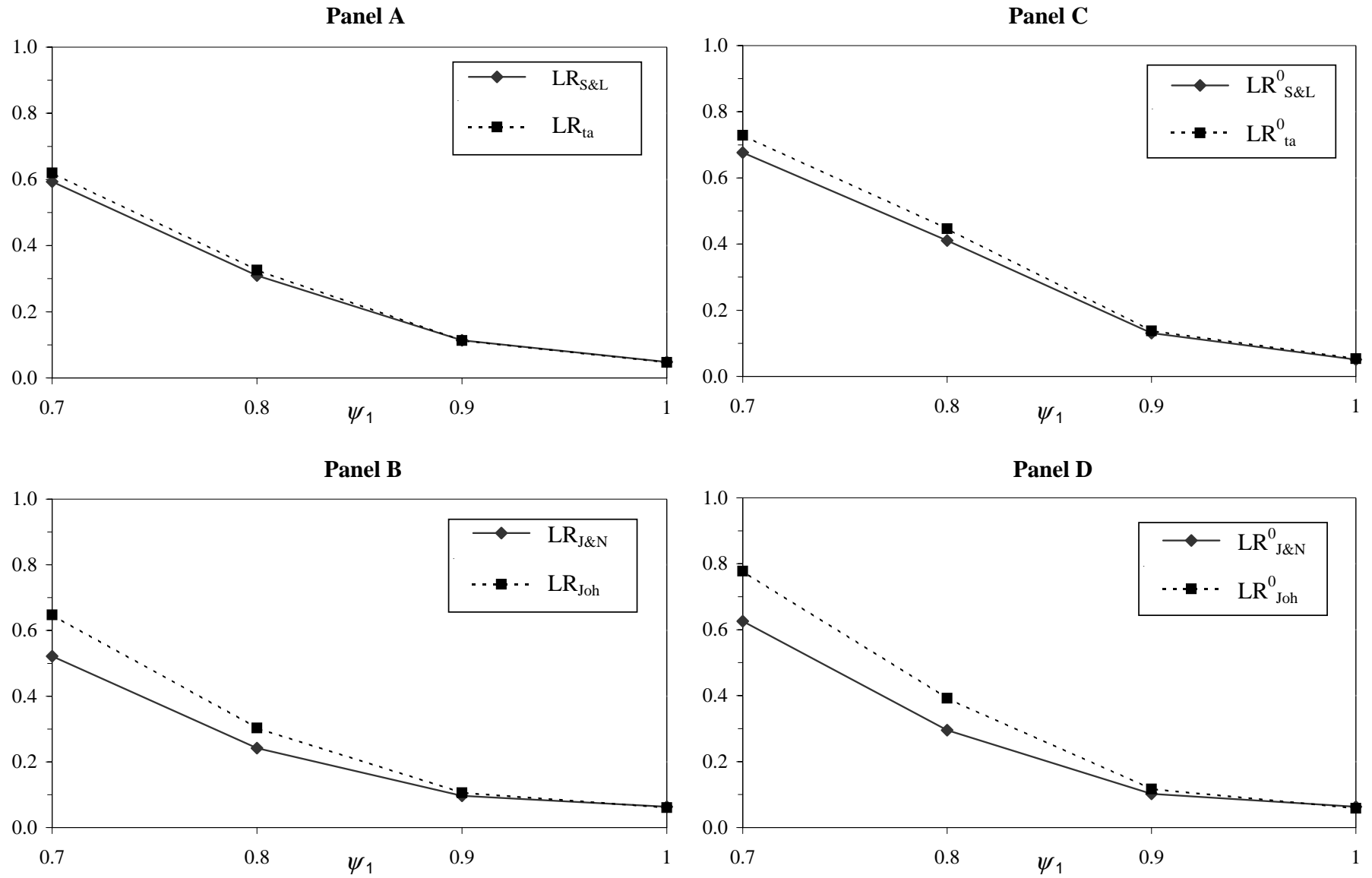


Figure 4. Small sample properties of tests for bivariate DGPs with $r = 0$ or $r = 1, r_0 = 0, T = 100, \theta = 0$

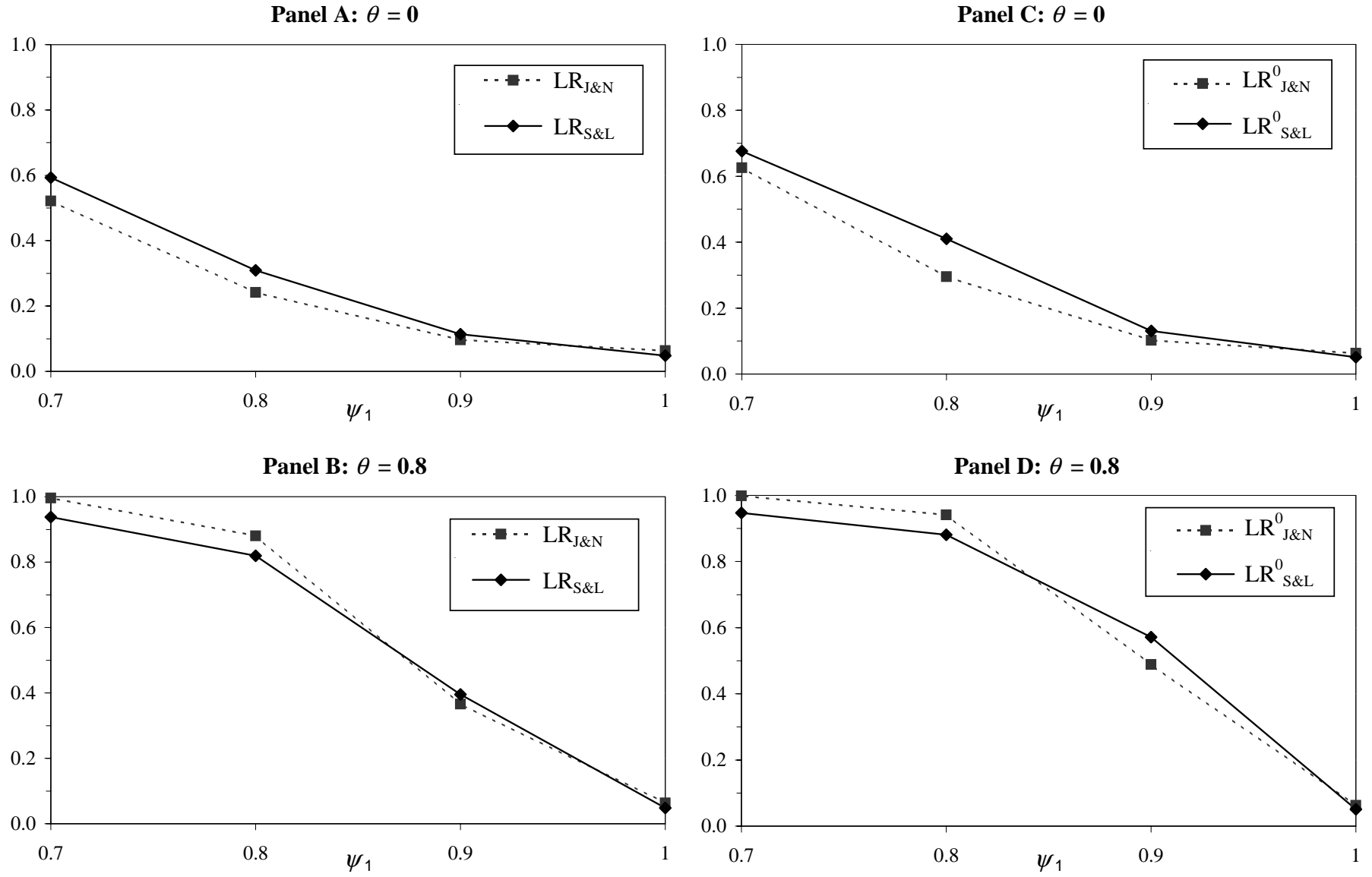


Figure 5. Small sample properties of tests for bivariate DGPs with $r = 0$ or $r = 1, r_0 = 0, T = 100$

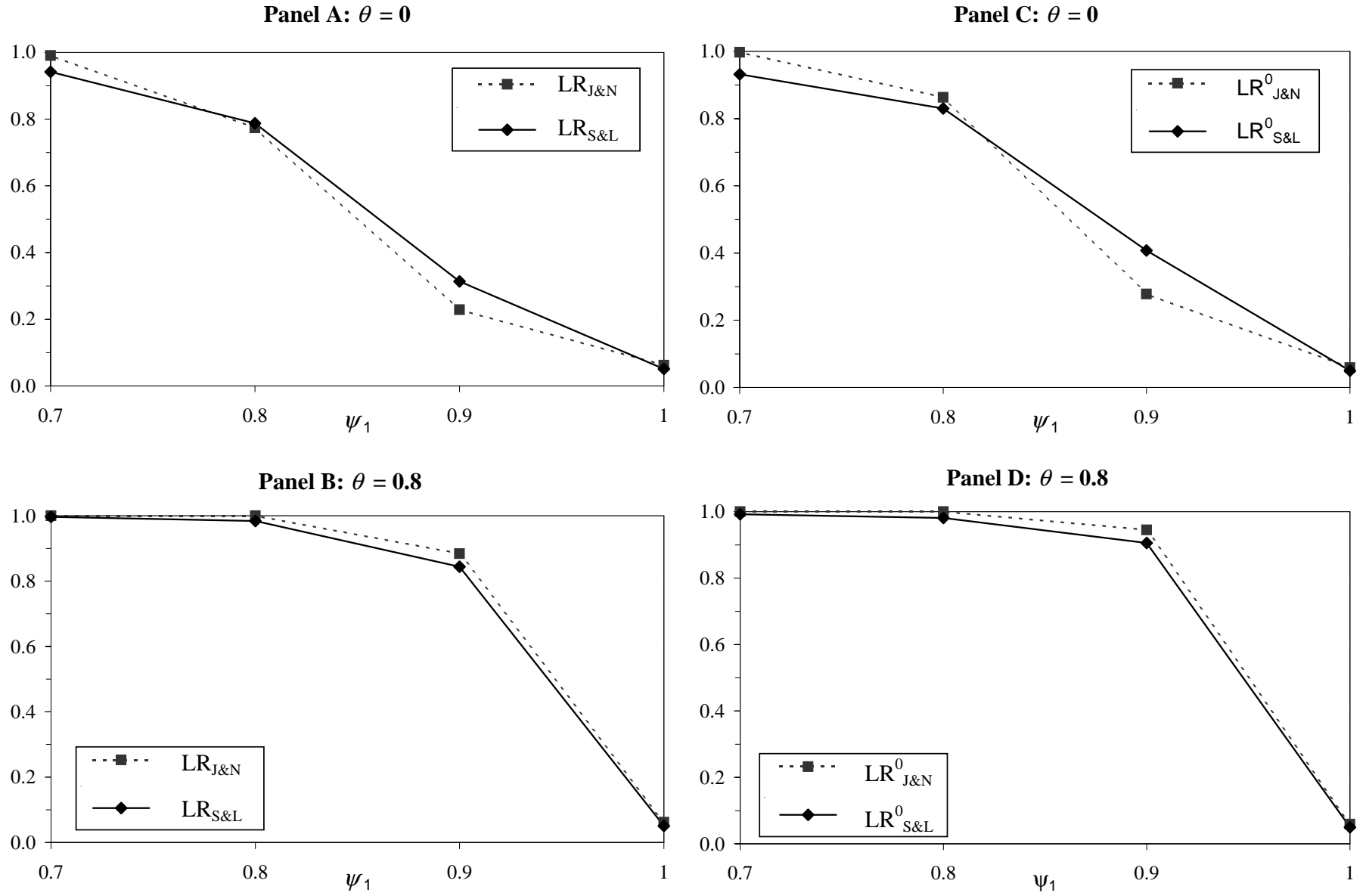


Figure 6. Small sample properties of tests for bivariate DGPs with $r = 0$ or $r = 1, r_0 = 0, T = 200$

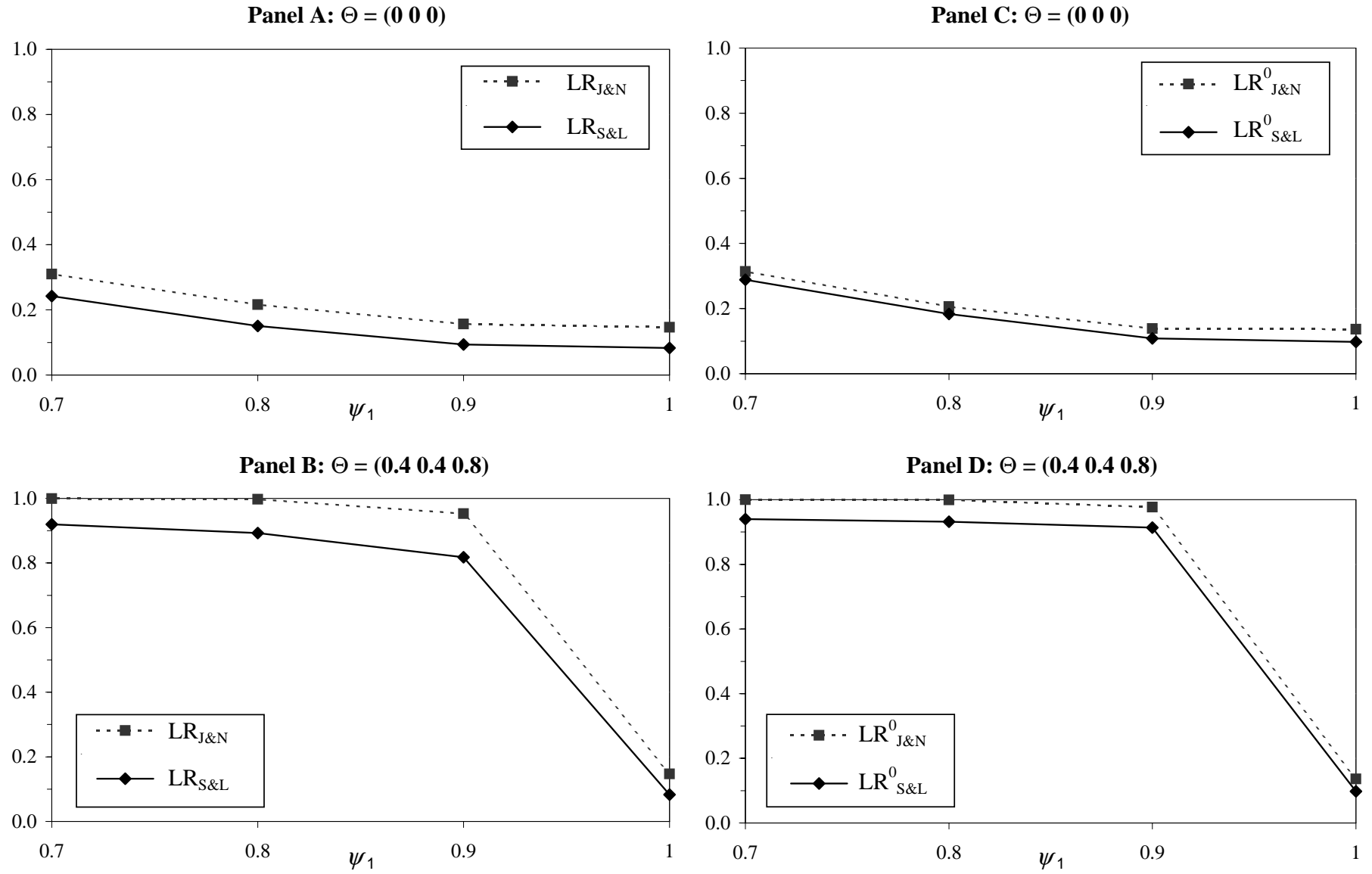


Figure 7. Small sample properties of tests for four-dimensional DGPs with $r = 0$ or $r = 1, r_0 = 0, T = 100$

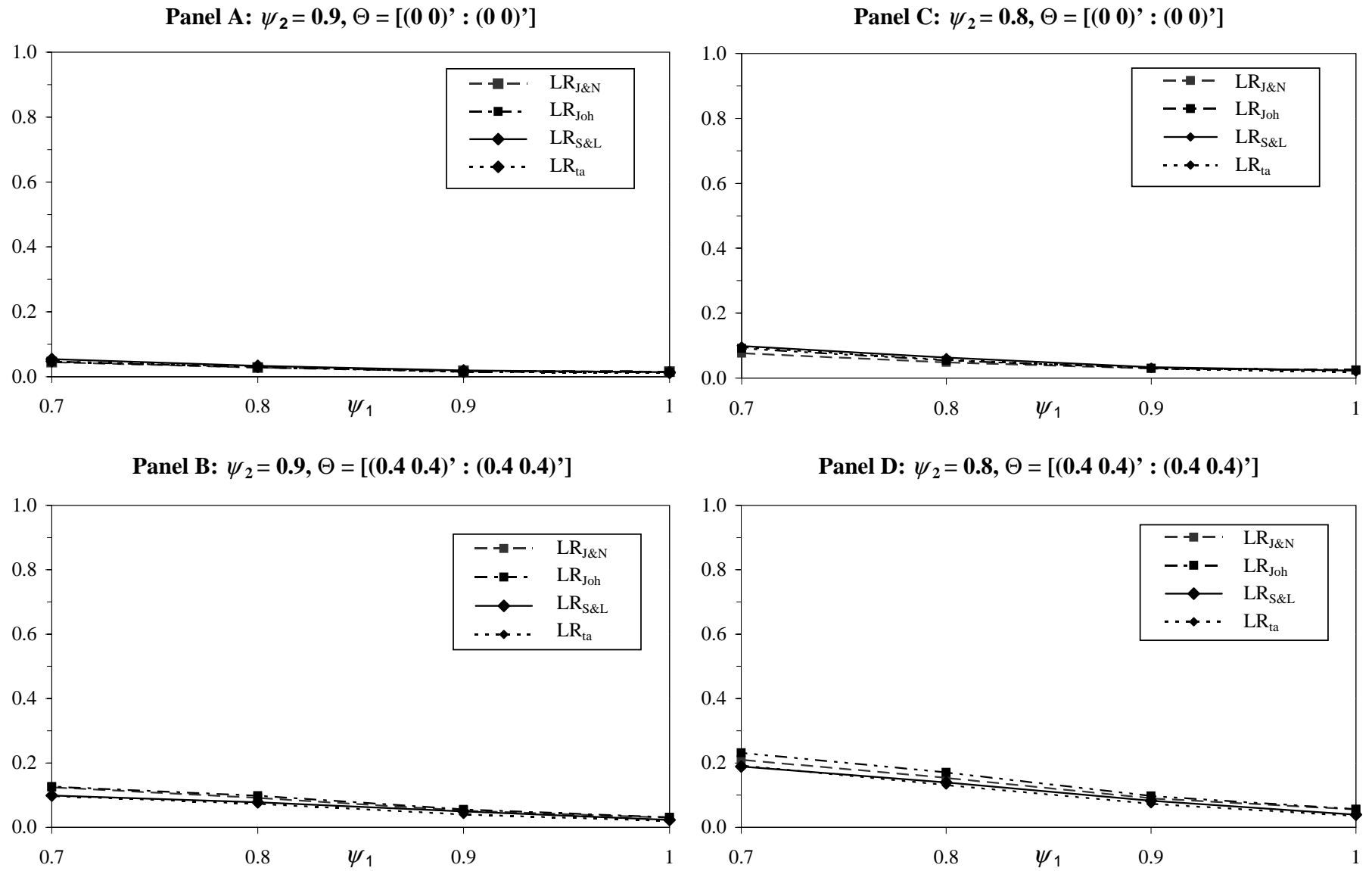


Figure 8. Small sample properties of tests for four-dimensional DGPs with $r = 1$ or $r = 2, r_0 = 1, T = 100$